Scaling in DNA unzipping models: denaturated loops and end-segments as branches of a block copolymer network

Marco Baiesi[1], Enrico Carlon[1,2], and Attilio L. Stella[1,3]
[1] *INFM - Dipartimento di Fisica, Università di Padova, I-35131 Padova, Italy*
[2] *Theoretische Physik, Universität des Saarlandes, D-66123 Saarbrücken, Germany*
[3] *Sezione INFN, Università di Padova, I-35131 Padova, Italy*
(February 6, 2008)

For a model of DNA denaturation, exponents describing the distributions of denaturated loops and unzipped end-segments are determined by exact enumeration and by Monte Carlo simulations in two and three dimensions. The loop distributions are consistent with first order thermal denaturation in both cases. Results for end-segments show a coexistence of two distinct power laws in the relative distributions, which is not foreseen by a recent approach in which DNA is treated as a homogeneous network of linear polymer segments. This unexpected feature, and the discrepancies with such an approach, are explained in terms of a refined scaling picture in which a precise distinction is made between network branches representing single stranded and effective double stranded segments.

PACS numbers: 05.20.-y, 05.70.Fh, 64.60.-i, 87.14.Gg

I. INTRODUCTION

A DNA molecule may undergo transitions from a double stranded to a single stranded state, either under the effect of an increase in temperature T (thermal denaturation), or through applied forces at one end of the chain (mechanical unzipping) [1–3]. In the characterization of such transitions, and in the determination of their universal, asymptotic features, substantial progresses were made recently by applications of models and methods of polymer statistics [4–8]. Among these progresses is an extension [6] of the classical Poland and Sheraga (PS) model [1]. In the PS model the partition function of a DNA chain is approximated by that of a sequence of non-interacting double-stranded segments and denaturated loops, and the thermal denaturation transition results of second order type [1,2]. Recently, excluded volume effects between a loop and the rest of the chain were included in the PS description in an approximate way [6], using results from the theory of polymer networks [9]. This approach predicts a first order denaturation, in agreement with very recent numerical studies of models taking fully into account the self and mutual avoidance among loops and double segments [7,8]. Quite remarkably, the approximate scheme of Ref. [6] yields results which are in good quantitative agreement with Monte Carlo simulations [8]. Most recently, predictions based on the theory of polymer networks were also made for the case of mechanical unzipping [10].

Besides confirming the expected first order character of thermal denaturation, the results of Ref. [8] demonstrated that excluded volume effects alone are responsible for this character, while the difference in stiffness between double and single stranded DNA, and sequence heterogeneity do not affect the asymptotic nature of the transition.

All these results rise interesting and debated [11] issues and open new perspectives in the field. First of all, one would like to test numerically the existing analytical estimates, in particular the new ones pertaining to the case of mechanical unzipping. The validity of crucial predictions relies on such tests, which should also reveal whether the quantitative success in the case of thermal unzipping is rather fortuitous, or there is some deep and systematic basis for it. The network picture proposed in Ref. [6], besides allowing some elegant and successful estimates, could constitute an important step forward in our way of representing the physics of DNA at denaturation. Progress in the assessment of the validity and limitations of a network picture for denaturating DNA should be allowed by a more careful and systematic analysis of numerical results for specific models. The possible extension and improvement of previous analyses for the relevant three dimensional case is also a main motivation of the present work.

In this article we consider a lattice model of DNA both in two and in three dimensions (d) with excluded volume effects fully implemented. By various numerical methods we estimate length distributions for denaturated loops and unzipped end-segments. In particular, for the transition in three dimensions we give here exponent estimates which extend and improve the results of a previous study [8]. Moreover, our results allow to address the basic issues mentioned above, and to clearly identify some qualitative and quantitative limitations of the picture proposed in Ref. [6] . The analysis gives also hints which allow us to propose a generalization of the polymer network representation of denaturating DNA. Our model in two dimensions, besides deserving some interest in connection with problems like the unzipping of double stranded polymers adsorbed on a substrate, offers an ideal context in which to compare the predictions of Refs. [6,10] with

numerical results. Indeed, for $d = 2$ those predictions are based on exactly known network exponents, while in $d = 3$ the same exponents have been approximately determined.

The model studied here was introduced in Ref. [7] and further analyzed and extended in Ref. [8]. We consider two self-avoiding walks (SAWs) of length N on square and cubic lattices, described by the vectors identifying the positions of each monomer $\vec{r}_1(i)$ and $\vec{r}_2(i)$, with $0 \le i \le N$. The walks, which represent the two DNA strands, have a common origin $\vec{r}_1(0) = \vec{r}_2(0)$ and only monomers with the same coordinate along the SAW's can overlap each other (i.e. $\vec{r}_1(j) = \vec{r}_2(k)$ only if $j = k$). An overlap corresponds to a bound state of complementary DNA base pairs, to which we assign an energy $\varepsilon = -1$, thus neglecting the effects of sequence heterogeneities. This is an acceptable approximation in view of the fact that, at coarse grained level, each monomer (site visited by the SAW) should represent a whole persistence length of the single strand, which includes several (≈ 10) bases. A denaturated loop of length l occurs whenever, for some i, $\vec{r}_2(i) = \vec{r}_1(i)$, $\vec{r}_2(i+l+1) = \vec{r}_1(i+l+1)$, and $\vec{r}_2(i+k) \neq \vec{r}_1(i+k)$, $k = 1, 2, .., l$.

At a temperature T each configuration ω of the two strands appears in the statistics with probability proportional to its Boltzmann's weight $\exp\{-\beta H(\omega)\}$, where $\beta = 1/T$ (Boltzmann's constant = 1) and H is ε times the number of bound base pairs in ω. The resulting behavior of the DNA model resembles the scaling of a $2N$-step SAW as long as the inverse temperature β is lower than a critical value β_c. For $\beta > \beta_c$ the strands are typically paired in a sequence of bound segments alternating with denatured loops, the latter gradually shrinking and becoming rarer for increasing β. For $\beta > \beta_c$ and $N \to \infty$ the scaling of an N-step SAW holds. Around β_c there is a crossover from one regime to the other, and exactly at β_c peculiar scaling behaviors are expected for the distributions of loops and end-segments.

II. THE TWO DIMENSIONAL CASE

We focus first on the two-dimensional case. Fig. 1 plots the logarithm of $P(l, N)$, the pdf of finding a denaturated loop of length l within a chain of total length N, as a function of $\ln l$ at the transition point. The behavior of this pdf at denaturation determines the order of the transition [1,6,8]. The data are obtained from exact enumerations for walks up to $N = 15$. For each chain length, the pdf is sampled at the temperature corresponding to the specific heat maximum. This choice offers the advantage that, if the specific heat diverges at the transition, like in our case, the correct transition temperature is asymptotically singled out for $N \to \infty$. Thus, in this limit, the sequence of distributions should automatically approach the critical pdf, for which we expect power-law scaling.

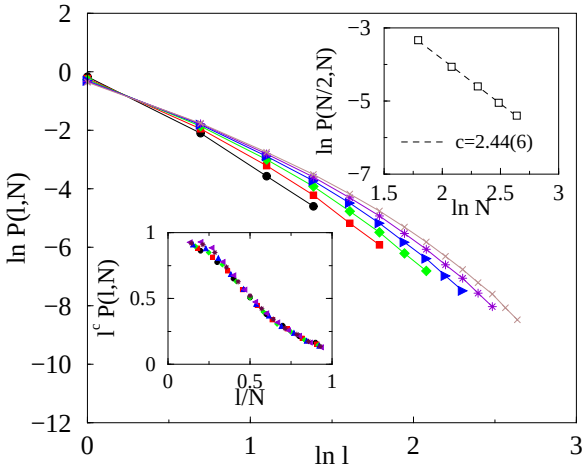

FIG. 1. $\ln P(l, N)$ vs. $\ln l$ for various N values. Upper inset: $\ln P(N/2, N)$ vs. $\ln N$; data are well-fitted by a line from which we estimate $c = 2.44(6)$. Lower inset: Scaling collapse of the data.

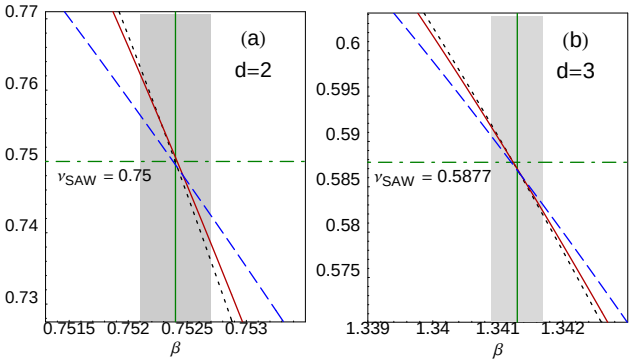

FIG. 2. (a) Effective exponent ν of the end-to-end distance, for $N \in A = \{80, 120, 160, 240\}$ (dashed curve), $N \in B = \{160, 240, 320, 480\}$ (dotted curve), and $N \in A \cup B$ (continuous curve), as a function of β in $d = 2$. The horizontal dotted line marks the exactly known SAW $\nu = 3/4$, while the vertical line is our estimate $\beta_c = 0.7525(3)$ (the error is indicated by the gray band). (b) Similar plots for $d = 3$ ($A = \{80, 120, 160, 240\}$, $B = \{120, 160, 240, 320\}$). In this case the intersections are even better localized around the expected SAW $\nu \approx 0.5877$. The transition temperature determined in this way is almost coincident with the estimate $\beta_c = 1.3413(4)$ [7] indicated by the vertical line and the gray band.

Finite-size effects should be described by $P(l, N) = l^{-c} q(l/N)$, with q a suitable scaling function. In order to determine c, we consider, e.g., $P(N/2, N)$, which should scale $\propto N^{-c}$ for $N \to \infty$. Such quantity is shown in the upper inset of Fig. 1, plotted as a function of N in a log-log scale. From a linear fit of the data we obtain the estimate $c = 2.44(6)$. The lower inset shows a scaling collapse of the data, obtained with $c = 2.44$. The good quality of the collapse indicates that the system is

very close to the asymptotic regime although the chains are quite short. We also performed a Monte Carlo determination of c using the pruned enriched Rosenbluth Method (PERM) [12], through which walks are generated by a growth procedure. In this case the critical temperature at which the pdf was sampled was determined by carefully monitoring the scaling with N of the average end-to-end distance $\langle|\vec{r}_2(N)-\vec{r}_1(N)|\rangle$ as a function of temperature. The effective exponents describing the growth with N of this quantity are plotted as a function of β for different chain lengths in Fig. 2 [13]. The intersections of the various curves in a very narrow range signals the crossover expected at denaturation for this quantity and allows to locate the melting temperature rather accurately, i.e. $\beta_c = 0.7525(3)$. By fitting the initial slope of the critical pdf for chains up to $N = 480$ we extrapolate $c = 2.46(9)$, in good agreement with the exact enumeration results.

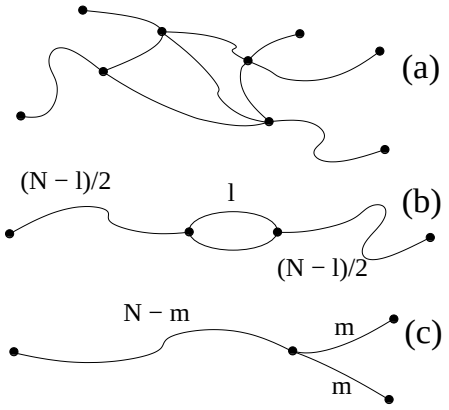

FIG. 3. (a) Configuration of a polymer network with $\mathcal{L} = 2$, $n_1 = 5$, $n_3 = 1$ and $n_4 = 3$. (b) Loop of length l embedded in a chain of length $N - l$ and (c) chain of length $N - m$ with bifurcating end segments of length m.

The value of c for loops with $l \ll N$ was predicted analytically [6] using exact results for entropic exponents of networks of arbitrary topology [9,14]. For a network of fixed topology G (see example in Fig. 3(a)), with n self- and mutually avoiding segments, Duplantier [9], on the basis of renormalization group arguments, postulated the following scaling form for the total number of configurations:

$$\Gamma_G \sim \mu^N N^{\gamma_G - 1} \quad f\left(\frac{l_1}{N}, \frac{l_2}{N}, .., \frac{l_n}{N}\right) \tag{1}$$

where l_i is the length of the i-th segment, $N = \sum_i l_i$ and f a scaling function. The value of γ_G depends on the number of independent loops, $\mathcal{L}$, and on the number of vertices with k legs, n_k, as

$$\gamma_G = 1 - \mathcal{L}d\nu + \sum_k n_k \sigma_k \tag{2}$$

where ν is the radius of gyration exponent and σ_k, $k = 1, 2, ..$, are exactly known exponents in $d = 2$ [9]:

$$\sigma_k = \frac{(2-k)(9k+2)}{64}\,. \tag{3}$$

This general scaling framework was applied to the DNA unzipping by considering relevant network topologies for the problem [6,10]. For example, in order to study the denaturated loop length pdf, one can assimilate the situation of a typical loop within DNA to that of the loop in Fig. 3(b). This amounts to assume that the action of the rest of the DNA molecule is the same as that of two long linear tails, thus totally disregarding the presence of other loops. According to Eq. (1), for G corresponding to the topology in Fig. 3(b), one has [6]

$$\Gamma_{\rm loop} \sim \mu^{2N} N^{-d\nu + 2\sigma_1 + 2\sigma_3} h(l/N)\,. \tag{4}$$

In this equation we explicitly assume that, while the connectivity of a loop step is equal to μ, that of a step of the tails is μ^2. This assumption is consistent with the thermodynamics of the denaturation transition and follows from the continuity of the canonical free energy of the system, and from the fact that each step of the tails corresponds in fact to two steps of the loops. In this way the l dependence of the r.h.s. of Eq. (4) does not enter in the exponential growth factor, an important requisite for the derivations below [15]. For $N \gg l$ the number of configurations should reduce to that of a single double stranded chain of length N, i.e. $\Gamma \sim \mu^{2N} N^{\gamma - 1}$ with $\gamma = 1 + 2\sigma_1$, according to Eq. (1). This requires $h(x) \sim x^{-2\nu + 2\sigma_3}$, for $x \ll 1$. So, the l dependence in Eq. (4) becomes $\sim l^{-c}$, with [6]

$$c = d\nu - 2\sigma_3\,. \tag{5}$$

Clearly, in this approximation, c is also the exponent by which the loop length pdf, P, scales. In reality, the segments departing from the two sides of the loop replace more complex fluctuating structures containing denaturated bubbles of all sizes, separated by short linear double stranded segments. In $d = 2$, $\nu = 3/4$ and $\sigma_3 = -29/64$, therefore $c = 2 + 13/32 \approx 2.41$, which is a value consistent within error bars with our numerical estimates. As already known in $d = 3$ [8], this agreement implies that the sequences of loops which "dress" the two segments departing from the loop in Fig. 3(b) have very little effects on the value of c.

We consider now the distribution of end-segments. With the assumed boundary conditions, denaturated end segments of length m occur in configurations where $\vec{r}_1(N-m) = \vec{r}_2(N-m)$ while $\vec{r}_1(k) \neq \vec{r}_2(k)$ for $k > N - m$. The statistical geometry of denaturated end-segments is expected to be relevant for situations occurring in mechanical unzipping experiments [10]. Indeed, as a rule, this unzipping is induced by applying forces which separate the strand extremes $\vec{r}_1(N)$ and $\vec{r}_2(N)$ by

micromanipulation techniques [16]. In the same spirit as in the case of denaturated loops, one can consider now the network geometry of Fig. 3(c) [10], for which

$$\Gamma_{\rm fork} \sim \mu^{2N} N^{3\sigma_1+\sigma_3} g(m/N)\,. \qquad (6)$$

By matching again, for $m \ll N$, with the partition function of a SAW of length N, one gets for the distribution of the end segment lengths $P_e(m,N) \sim m^{-\bar{c}}$ with [10]

$$\bar{c} = -\left(\sigma_1 + \sigma_3\right)\,. \qquad (7)$$

One should note that this last result holds also in case one tries to match the behavior of $\Gamma_{\rm fork}$ with that of a simple SAW of length $2N$ in the limit $(N-m) \ll N$. In this case the pdf results $\sim (N-m)^{-\bar{c}}$.

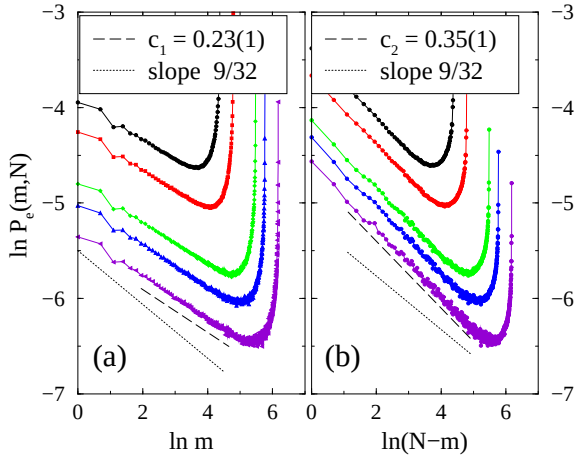

FIG. 4. Plot of $\ln P_e(m,N)$ vs. $\ln m$ (a) and $\ln(N-m)$(b) at the estimated critical point $\beta_c = 0.7525$ and for $N = 80$, 120, 240, 320 and 480 (from top to bottom).

Figure 4 shows a plot of the logarithm of the pdf for end segments of length m, $\ln P_e(m,N)$, as a function of (a) $\ln m$ and (b) $\ln(N-m)$. The data obtained from the PERM show coexistence of two distinct power-law scaling behaviors, i.e. $P_e(m,N) \sim m^{-c_1}$ and $P_e(m,N) \sim (N-m)^{-c_2}$ where a linear fit yields $c_1 = 0.23(1)$ and $c_2 = 0.35(1)$, respectively. The slope determinations are rather sharp in this case, and, in fact, the criterion of selecting the denaturation temperature in correspondence with the simultaneous manifestation of the power law behaviors of P_e is very efficient, and consistent with that based on monitoring the behavior of $\langle|\vec{r}_2(N) - \vec{r}_1(N)|\rangle$.

The existence of two distinct slopes, and the values of the exponents are in disagreement with what one expects on the basis of the network approximation, i.e. $\bar{c} = 9/32 \approx 0.28$ in $d = 2$. Thus, for some reason, on $P_e(m,N)$ the effects of the structure of the double stranded part of the chain are noticeable and the schematization through a simple polymer network topology is not fully adequate to represent the physics. In the case $N - m \ll N$ the number of single chain configurations is indeed that of a simple linear SAW of length $2N$, which is asymptotically exactly known in two dimensions [4]. On the other hand, for $m \ll N$, the configurations to count are those of an effective linear SAW chain of length N, whose internal structure contains loops at all scales. It is conceivable that not only the connective constants, but also the entropic scaling properties of such effective walk differ from those of a standard SAW. The difference must concern the power law correction to the exponential growth factor of the number of configurations as a function of N. Indeed, as already remarked above, the connectivity constant of the effective, double stranded walk at the transition, must be just the square of the simple SAW connectivity constant. This follows from an obvious requirement of free energy continuity and from the fact that in the high temperature region the two strands behave as unbound simple SAW, with a total length twice that of the effective double stranded walk, independent of temperature [17].

To investigate this issue further we determined directly, on the basis of PERM data, also the overall entropic behavior of the DNA chain. We considered two types of boundary conditions: (1) the extremes ($\vec{r}_1(N)$ and $\vec{r}_2(N)$) of the chain are free and (2) are forced to join in a single point ($\vec{r}_1(N) = \vec{r}_2(N)$), while in both cases, the strands have still a common origin ($\vec{r}_1(0) = \vec{r}_2(0)$).

Condition (1) is the one applying to the effective walk discussed above. We indicate with $Z_N^{(1)}$ and $Z_N^{(2)}$ the corresponding partition functions. In the spirit of the network approximation, we neglect the contribution of denaturated loops within the double stranded phase one has $Z_N^{(2)} \sim \mu^{2N} N^{\gamma^{(2)}-1}$, with $\gamma^{(2)} = 1 + 2\sigma_1$. We estimate also $Z_N^{(1)}$ within the same general framework, by integrating the partition function $\Gamma_{\rm fork}$ of Eq. (6) over all possible end-segment lengths. This integration gives

$$Z_N^{(1)} \sim \int_0^N dm\ \ \Gamma_{\rm fork} \sim \mu^{2N} N^{3\sigma_1+\sigma_3+1}, \qquad (8)$$

as the integration of the scaling function $g(m/N)$ yields an extra factor proportional to N. Defining $Z_N^{(1)} \sim \mu^{2N} N^{\gamma^{(1)}-1}$ one eventually gets:

$$\gamma^{(1)} = 2 + 3\sigma_1 + \sigma_3\,. \qquad (9)$$

This is a theoretical expression for the entropic exponent of a DNA molecule as a whole, and can be directly compared with numerical estimates, one of which already exists in $d = 3$ [7], as we discuss in the next sections.

To calculate the entropic exponents $\gamma^{(1)}$ and $\gamma^{(2)}$ we estimated the quantity $(Z_{2N}^{(i)}/Z_N^{(i)})^{1/2N}$ by PERM sampling for reasonably long chains. For $N \to \infty$ one expects

$$\left(\frac{Z_{2N}^{(i)}}{Z_N^{(i)}}\right)^{\frac{1}{2N}} \sim \mu\left(1 + \ln 2\ \frac{\gamma^{(i)}-1}{2N}\right), \qquad i = 1,\,2\,. \qquad (10)$$

Since μ must coincide with the SAW connective constant, which in $d = 2$ is very precisely known ($\mu = 2.63815852927(1)$ [18]), large N data for the quantity on the l.h.s. of Eq. (10) can be fitted by keeping as unique fitting parameter $\gamma^{(i)}$ in the r.h.s. expression. Alternatively, one can assume a value for $\gamma^{(i)}$ and check whether data appear consistent with the assumed correction term in the same r.h.s. expression. This consistency test is best applied to sets of data pertaining to different temperatures close to the transition, as illustrated in the next section. Since the correction term is rather sensitive to the choice of the temperature, this offers also a way of locating the transition. Here we just mention that the slopes predicted on the basis of the network approximation in $d = 2$ appear slightly, but definitely, inconsistent with the data, as discussed in the next section.

III. DENATURATING DNA AS A COPOLYMER NETWORK

The polymer network representation of denaturating DNA [6] was originally proposed as a useful, but definitely approximate tool, without the possibility of controllable and systematic improvements. For example, the environment seen by a single fluctuating loop within the molecule was not proposed as something unique, and choices slightly different from that discussed above were also discussed [6,10]. With these alternative choices, leading to slightly different results, the portions of the molecule surrounding the loop were not necessarily treated as made of simple double segments.

Even if the approximate character remains, all the discrepancies and inconsistencies discussed in the previous section can be resolved by a refinement of the whole picture and an improvement of the approach, which could allow more accurate predictions in the future. In the new perspective we propose, the rules for associating a network schematization to loops or end-segments should be regarded as unique. This follows from the fact that the new polymeric entities one defines are supposed to account, at an effective level, for all the complications arising from the fluctuating geometry of the model.

The legs of the network are either simple self-avoiding chains, in case they are really made of single strands, or effective, dressed segments, when they represent fluctuating double stranded portions of the molecule comprised between two given bound base pairs. This clear cut distinction suggests that the exponents σ_k, associated with the network vertices, should be modified with respect to the "bare" values considered so far, as soon as at least one outgoing leg is of the double stranded, dressed type. Indeed, within the framework of a continuum Edwards model description of the inhomogeneous polymer network, and of renormalization group ideas, one can postulate the validity of homogeneity laws and exponent relations analogous to Eqs. (1) and (2), with modified σ's where appropriate. Something like this was already done in Ref. [19] where explicit field theoretical calculations were performed in the case of copolymer star networks (no loops). These stars were composed of mutually excluding branches made either of random walks, or of self avoiding walks. In the case considered here the copolymer is made of two kinds of segments which do not differ as far as metric scaling exponents are concerned (same ν), but can have different entropic scaling properties. It is indeed quite conceivable that an effective linear structure, which should be resolved into a sequence of denaturated segments and loops of all sizes, could have an entropic scaling different from the one of a simple SAW on the lattice, even if the elongation grows in the same way with N.

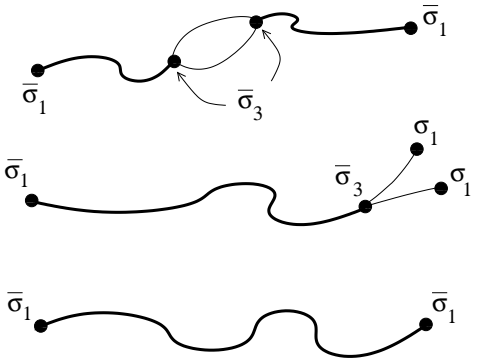

FIG. 5. Representation of the geometries relevant for the DNA problem by copolymer networks. Thick lines denote branches of the network composed by an alternating sequence of bound segments and denaturated loops, while thin lines are genuine single stranded segments. Only the latter have the same metric and entropic scaling properties as a SAW.

The above considerations lead us to introduce two new exponents: $\overline{\sigma}_1 = \sigma_1 + \Delta\sigma_1$ for an isolated vertex with a dressed segment, and $\overline{\sigma}_3 = \sigma_3 + \Delta\sigma_3$ for a vertex joining a dressed segment and two self avoiding walks, as schematically illustrated in Fig. 5. As discussed above, the introduction of such $\overline{\sigma}$'s amounts to postulating a generalization of the scaling in Eq. (1) to copolymer networks with two different types of segments. Following the same arguments leading to Eqs.(5), (7) and (9) and using now the modified exponents wherever appropriate, one finds:

$$c_1 = -(\sigma_1 + \sigma_3) - \Delta\sigma_3 + \Delta\sigma_1 \ , \tag{11}$$

$$c_2 = -(\sigma_1 + \sigma_3) - \Delta\sigma_3 - \Delta\sigma_1 \ , \tag{12}$$

$$c = d\nu - 2\sigma_3 - 2\Delta\sigma_3 \ , \tag{13}$$

$$\gamma^{(2)} = 1 + 2\sigma_1 + 2\Delta\sigma_1 \ , \tag{14}$$

$$\gamma^{(1)} = 2 + 3\sigma_1 + \sigma_3 + \Delta\sigma_1 + \Delta\sigma_3 \ , \tag{15}$$

which is a set of five equations with only two unknown parameters, $\Delta\sigma_1$ and $\Delta\sigma_2$. These equations should be

regarded as consistency requirements in order to test the validity of the proposed copolymer picture. Notice that within the copolymer network scheme the exponents associated to the distributions of short and long end-segments are distinct, $c_1 \neq c_2$, as soon as $\Delta\sigma_1 \neq 0$. By solving the first two equations (11) and (12) with the numerical values for c_1 and c_2, we find $\Delta\sigma_3 = -0.01(1)$ and $\Delta\sigma_1 = -0.06(1)$. While the former is actually compatible with zero, the latter is not. Once the values of $\Delta\sigma_1$ and $\Delta\sigma_3$ have been fixed, we can check for the consistency of the other exponents using Eqs. (13), (14) and (15). Inserting the calculated $\Delta\sigma_3$ into Eq. (13) we find $c = 2.43(2)$, i.e. a value slightly higher than what predicted from Eq. (5). Indeed, our numerical estimates suggest for c a slightly higher value than that predicted on the basis of the bare network approximation, although the error bars cover both values. Next we consider the last two equations (14) and (15); as both $\Delta\sigma_1$ and $\Delta\sigma_3$ are negative we expect that $\gamma^{(1)}$ and $\gamma^{(2)}$ would be somewhat smaller than the values predicted from the homogeneous network approximation. Substituting the above numerical values of $\Delta\sigma_1$ and $\Delta\sigma_3$ into Eqs. (14) and (15) we find $\gamma^{(2)} \approx 1.22(2)$ and $\gamma^{(1)} \approx 1.99(2)$ (recall that the homogeneous network predictions are $\gamma^{(2)} = 1 + 11/32 \approx 1.34$ and $\gamma^{(1)} = 2 + 1/16 \approx 2.06$). We first examine case (1). Since the differences between the "bare" and "dressed" values of $\gamma^{(1)}$ are small, to magnify the asymptotic details we considered the quantity

$$f(N) \equiv \left(\frac{Z_{2N}^{(1)}}{Z_N^{(1)}}\right)^{\frac{1}{2N}} - \frac{(\gamma^* - 1)\mu \ln 2}{2N} , \qquad (16)$$

with $\gamma^* = 33/16$. The coefficient of the $1/N$ term in the r.h.s. of the previous expression has been chosen such that if $Z_N^{(1)}$ would scale with the homogeneous network exponent then $f(N)$ should approach the connectivity constant μ with zero slope when plotted as function of $1/2N$ (as follows from Eq. (10)). A $\gamma^{(1)}$ larger (smaller) than its homogeneous network value would imply a $f(N)$ approaching μ with a positive (negative) slope, equal to $(\gamma^{(1)} - \gamma^*)\mu \ln 2$. Figure 6 shows a plot of $f(N)$ vs. $1/2N$ for five different temperatures around the estimated critical one. Clearly $f(N)$ approaches μ in the limit of large N with a negative slope. The solid line represents the slope for the value of $\gamma^{(1)} = 1.99$ as obtained from Eq. (15), which apparently fits reasonably well the data.

In the case of $Z_N^{(2)}$, from all configurations generated by the PERM only those where the two strands have common endpoints ($\vec{r}_1(N) = \vec{r}_2(N)$) are considered and this reduces considerably the statistics available. The numerical results we find are consistent with both proposed pictures, since the low precision of data, due to the insufficient sampling, produces large statistical errors. So, the case (2), while also giving meaningful data, does not help in the determination of the right scenario.

We conclude that the improved representation based on the idea of a copolymer network with modified entropic exponents allows to match well all our numerical results, with just two adjustable parameters.

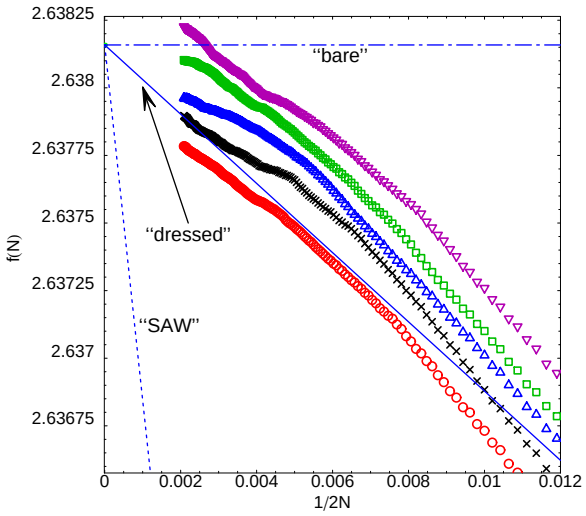

FIG. 6. The function $f(N)$, defined in Eq. (16), for $N \to \infty$ should approach μ with zero slope if the $\gamma^{(1)}$ exponent would be correctly determined by the homogeneous network approximation. Instead, the trend predicted by the copolymer theory (continuous line) fits rather well the data ($\beta = 0.7520, 0.7523, 0.7526, 0.7529, 0.7532$, from below. We recall that we estimate $\beta_c = 0.7525(3)$).

IV. THE THREE DIMENSIONAL CASE

We now consider the $d = 3$ case, where homogeneous network exponents can be deduced from estimates of $\gamma = 1 + 2\sigma_1$ [20] ($\sigma_1 \approx 0.079$) or from ε-expansion results combined with resummation techniques ($\sigma_3 \approx -0.175$) [21]. From Eqs.(5) and (7) one gets $c \approx 2.11$ and $\bar{c} \approx 0.095$. Previous Monte Carlo simulations yielded $c = 2.10(4)$ [8]. In the present work we made an extra effort in order to get a reliable estimate of the various exponents. With an analysis of the scaling behavior of $\langle|\vec{r}_1(N) - \vec{r}_2(N)|\rangle$ we get here a very precise estimate of the melting temperature [see Fig. 2(b)], consistent with that of Ref. [7]. Extensive PERM sampling of the loop distribution at the estimated transition temperature yields $c = 2.18(6)$ (see Fig. 7). This refined value, confirming the first order character of the transition, is slightly higher than that reported in Ref. [8], although compatible within the uncertainties. The discrepancy could be imputed to a slight overestimation of the transition temperature made in that reference. An analysis of the end-segments distribution yields, as in $d = 2$, two slightly different exponents, $c_2 > c_1$, signalling again deviation from the prediction of the homogeneous network approximation. We estimate $c_2 = 0.16(1)$ and

$c_1 = 0.14(1)$. These estimates are rather sharp, and imply $\Delta\sigma_1 = 0.01(1)$ and $\Delta\sigma_3 = 0.055(10)$. By inserting these values for the $\Delta\sigma$'s into our expression for c, Eq. (11), we get $c \approx 2.22$, which is well compatible with the last mentioned estimate illustrated in Fig. 7.

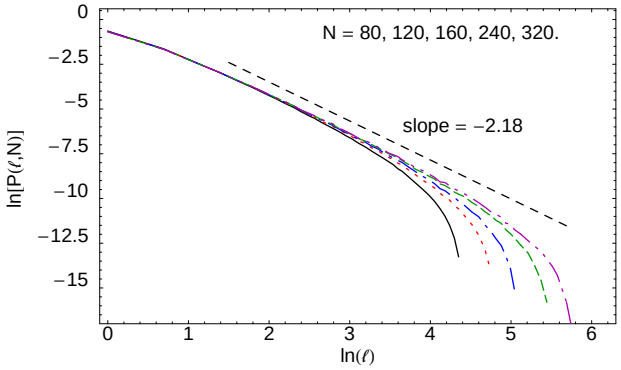

FIG. 7. Log-log plot of the loops pdf at the critical point as function of their length for chains of various lengths.

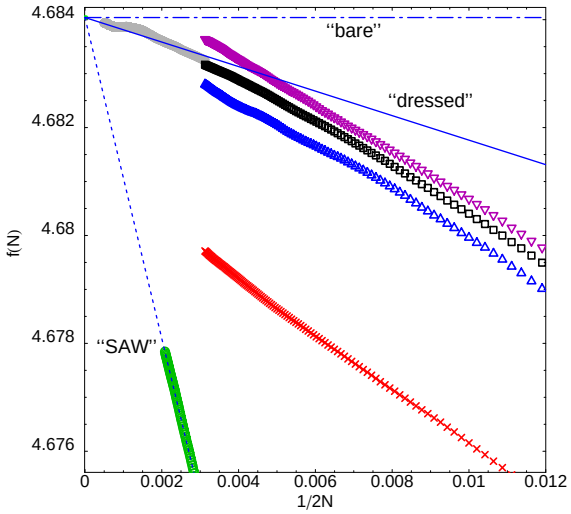

FIG. 8. As in Fig. 6: from below, $\beta = 0.293$ (SAW regime, evidenced by the dotted line, i.e. $\gamma^* = \gamma$, with gamma representing the SAW exponent in $d = 3$ [20]), 1.335, 1.3407, 1.3413(β_c), 1.3419. The gray band shows data at β_c and longer N (up to 2000), but with lower statistics.

Causo *et al.* [7] determined in $d = 3$ the exponent $\gamma^{(1)} = 2.09(10)$, a value which should be compared with the Eqs. (9) and (15) derived in this paper. Inserting the above determined values for $\Delta\sigma_1$ and $\Delta\sigma_3$ in Eq. (15) we obtain $\gamma^{(1)} \approx 2.00$, to be compared with a value $\gamma^{(1)} \approx 2.06$ if one sticks to the homogeneous network prediction, given by Eq. (9). As the $\gamma^{(1)}$ of Ref. [7] is compatible with both above values we made an effort to estimate it again with improved accuracy.

Figure 8 shows a plot of $f(N)$ vs. $1/2N$ for five different temperatures of which three around β_c (as for the $d = 2$ case, we set γ^* equal to the value derived within the homogeneous network picture for $\gamma^{(1)}$, $\gamma^* = 2.06$). The value of the connectivity constant is $\mu = 4.68404(9)$ for the cubic lattice [22], a value indeed approached by all data sets for $\beta \leq \beta_c$. At high temperatures the data clearly show SAW scaling, as expected, while at the transition point $f(N)$ seems indeed to approach a slope given by the "dressed" exponent $\gamma^{(1)} = 2.00$ (solid line). To confirm this, at the estimated critical point, we performed a series of calculations up to very long chains ($N = 2000$), but with somewhat lower statistics. The latter data are plotted in gray in Fig. 8: indeed they seem to follow quite closely the solid line, as predicted by the inhomogeneous network theory.

Thus, even if the error bars on some exponents are still relatively large, the corrections suggested by a block copolymer network representation of DNA seem to be well consistent with the numerical scenario, like in the two-dimensional case.

V. CONCLUSION

We investigated a lattice model of DNA denaturation both in $d = 2$ and $d = 3$ and determined the exponents associated to the decays of the pdf's for loops and end-segments. For the loop pdf exponent we find $c > 2$, which implies a first order denaturation as the average loop length remains finite at the transition point [1,2]. In $d = 2$ the first order character is even more pronounced than in $d = 3$ and the values of c are quite consistent with those predicted analytically on the basis of entropic exponent of homogeneous networks [6]. An analysis of the end-segment lengths reveals that the corresponding pdf displays two distinct power law behaviors, one applying at small, and the other at large fork openings. This unexpected feature, not predicted within the framework of the approximate description based on homogeneous polymer networks [6], suggests to describe the DNA fluctuating geometry as a copolymer network, in which a distinction is made between single stranded and effective double stranded segments. The latter are assumed to have different entropic exponents than SAW's, which lead to the introduction of a generalization of the previously known expression for pdf and entropic exponents (see Eqs. (11-15)). We tested the compatibility of the observed results with this effective copolymer network picture, which is assumed to catch the essential physics of the transition. Within this new framework, the different scaling behaviors of the end-segment pdf can be qualitatively explained, and the notion of overall entropic scaling of the macromolecule acquires a precise and consistent meaning. Especially in $d = 2$, the numerical evidence that a block copolymer network picture is compatible with the overall data, is rather clear and convincing.

A main reason why the copolymer network description turns out to be well compatible with the observed scalings is probably the fact that the pdf of loop length scales always with a sufficiently large c exponent. So, even for $N \to \infty$, the average width of the loops, and thus also of the dressed segments, remains finite. This could be an important requisite for the validity of the network picture. Indeed, it would be interesting to test whether similar copolymer pictures work also for other unzipping transitions of double stranded polymers with a smaller c [10]. A natural candidate is the unzipping occurring for the diblock copolymer model studied in Ref. [23]. The physics of that system can be assimilated to that of a DNA molecule in which each of the two strands is made exclusively of one type of base, and the bases of the two strands are complementary [10]. Of course, in such a model the loops forming can be made with portions of different lengths of the two strands. Moreover, different loops can also bind and form more complicated topologies than in the DNA case. By applying the simple homogeneous network picture to such a model, one would expect a second order transition with [10] $c = d\nu - 2\sigma_3 - 1 \approx 1.4$ in $d = 2$. The results for the copolymer model of Ref. [23], suggest a slightly, but definitely higher value $c = 1 + 9/16 \approx 1.56$, which is supported by a connection with percolation theory. This discrepancy could be due to the more complicated topology of the loops and could indicate that the model is less favorable for the application of a network picture.

Homogeneous networks are very interesting *per se* and most recently were recognized as important tools also for the study of the topological entanglement of polymers [24]. Block copolymer networks are a still relatively unexplored subject, in spite of the obvious fundamental and applicative interest. Thus, the realization that such networks could also be very relevant for a description of system like DNA at denaturation, at the moment, can not yield quantitative predictions based on field theory results. Indeed, even the problem of identifying which kind of copolymer model in the continuum, if any, could represent correctly our discrete DNA in the scaling limit is far from trivial and completely open. For sure the identification of DNA as a system potentially connected to the copolymer network physics adds further interest to these intriguing objects, which are already known to be somehow related to multifractal aspects of polymer statistics [19] and, in the $d = 2$ case, could realize interesting examples of conformal invariance.

We thank D. Mukamel for fruitful discussions and E. Orlandini for ongoing collaboration. Financial support by MURST through COFIN 2001, INFM through PAIS 2001 and by European Network on "Fractal and Structures and Self-Organization" are gratefully acnkowledged.

[1] D. Poland and H. A. Sheraga, J. Chem. Phys. **45**, 1456 (1966); **45**, 1464 (1966).
[2] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[3] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).
[4] See, e.g., C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge 1998).
[5] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989); D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997); S. M. Battacharjee, J. Phys. A **33**, L423 (2000); E. Orlandini, S. M. Bhattacharjee, D. Marenduzzo, A. Maritan, and F. Seno J. Phys. A **34**, L751 (2001); D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000); D. Marenduzzo, S. M. Bhattacharjee, A. Maritan, E. Orlandini, and F. Seno, Phys. Rev. Lett. **88**, 028102 (2002).
[6] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[7] M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000).
[8] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).
[9] B. Duplantier, Phys. Rev. Lett. **57**, 941 (1986).
[10] Y. Kafri, D. Mukamel, and L. Peliti, cond-mat/0108323.
[11] A. Hanke and R. Metzler, cond-mat/0110164; Y. Kafri, D. Mukamel, and L. Peliti, cond-mat/0112179.
[12] P. Grassberger, Phys. Rev. E 56, 3682 (1997).
[13] The data are extrapolated at temperatures different from the ones used in PERM runs by using the multiple histogram method: A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1195 (1989).
[14] K. Ohno and K. Binder, J. Phys. France **49**, 1329 (1988).
[15] One should notice that already at this stage the necessity of attributing different connective constants to different segments confers a block-copolymer character to the network. We will see below that the copolymer character should naturally involve a more essential distinction of universal entropic exponents, in addition to that of connective constants.
[16] A. Ashkin, Proc. Natl. Acad. Sci. USA, **94**, 4853 (1997).
[17] For a simpler model of copolymer unzipping this kind of property has been recently shown exactly: P. Leoni, C. Vanderzande, and L. Vandeuren, J. Phys. A **34**, 9777 (2001).
[18] I. Jensen and A. J. Guttmann, J. Phys. A **32**, 4867 (1999).
[19] C. von Ferber and Yu. Holovatch, Phys. Rev. E **56** 6370 (1997).
[20] S. Caracciolo, M. S. Causo, A. Pelissetto, Phys. Rev. E **57**, R1215 (1998).
[21] L. Schäfer, C. von Ferber, U. Lehr, and B. Duplantier, Nucl. Phys. **B374**, 581 (1989).
[22] D. MacDonald *et al.*, J. Phys. A **33**, 5973 (2000).
[23] E. Orlandini, F. Seno, and A. L. Stella, Phys. Rev. Lett. **84**, 294 (2000); M. Baiesi, E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. E **63**, 041801 (2001).
[24] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar, cond-mat/0202075.